CHILD DEVELOPMENT

Child Development, May/June 2021, Volume 92, Number 3, Pages e252–e269

Modeling and Visualizing the Codevelopment of Word and Nonword Reading in Children From First Through Fourth Grade: Informing Developmental Trajectories of Children With Dyslexia

Laura M. Steacy and Ashley A. Edwards
Florida Center for Reading Research, Florida State University

Jay G. Rueckl
University of Connecticut

Yaacov Petscher and Donald L. Compton
Florida Center for Reading Research, Florida State University

Developmental studies examining relations between word reading (WR) and decoding in typical and dyslexic populations routinely cut the reading distribution to form distinct groups. However, dichotomizing continuous variables to study development is problematic for multiple reasons. Instead, we modeled and visualized the parallel growth of WR and nonword reading (NWR) factor scores longitudinally in a Grade 1–4 developmental sample (N = 588). The results indicate that while WR and NWR growth factors are highly related (r = .71), the relation between WR and NWR trajectories change as a function of initial WR. Results are interpreted within computational models of dyslexia in which children with dyslexia overfit orthography → phonology relations at the level of the word, limiting the development of sublexical representations needed to read nonwords.

Dyslexia is a neurodevelopmental disorder characterized by difficulties with accurate and/or fluent word recognition and by poor spelling and decoding abilities that are not accounted for by specific sensory deficits or more general intellectual impairment (Fletcher, 2009). Early and prolonged difficulties in the acquisition of context-free and automatic word identification (WID) and decoding skills are the most reliable indicators of developmental dyslexia (Lovett et al., 1994; Torgesen, 2000) with deficits persisting across the life span (see Bruck, 1993; Shaywitz et al., 1999). The neurocognitive influences on dyslexia are multifactorial and primarily involve phonological processing deficits as well as weaknesses in other oral language skills and processing speed (Peterson & Pennington, 2015). Phonological processing deficits—impaired representation of, or access to, the abstract units of spoken language—have been implicated as the principal source of reading difficulties in dyslexic children by disrupting the ability to attend to and manipulate linguistic sounds crucial to the establishment and automatization of letter–sound correspondences (known as sublexical orthographic—phonological relations), which in turn underlie accurate and fluent word recognition development through the process of phonological decoding (Pennington, 2009); known as self-teaching (Share, 1995). As such, deficits in individuals with dyslexia are more likely to be observed when the phonological demands of the word reading (WR) task are greater (e.g., decoding of pronounceable nonwords), giving rise to the well-documented nonword reading (NWR) deficit in children with developmental dyslexia (e.g., Harm & Seidenberg, 1999; Metsala, Stanovich, & Brown, 1998; Rack, Snowling, & Olson, 1992).

Epidemiological studies have shown that dyslexia is a dimensional disorder and exists at the lower end of the normal continuum of WR ability (Shaywitz, Escobar, Shaywitz, Fletcher, & Makuch, 1992). In dimensional disorders, determining where on the continuum (in this case WR and decoding

This research was supported in part by Grant R324G060036 from the Institute of Education Sciences, U.S. Department of Education, and Grant P20HD091013 from the Eunice Kennedy Shriver National Institute of Child Health and Human (NICHD). Statements do not reflect the position or policy of these agencies, and no official endorsement by them should be inferred.

Correspondence concerning this article should be addressed to Laura M. Steacy, Florida Center for Reading Research, Florida State University, 2010 Levy Avenue, Suite 100, Tallahassee, FL 32310. Electronic mail may be sent to lsteacy@fcrr.org.

© 2020 Society for Research in Child Development
All rights reserved. 0009-3920/2021/9203-0029
DOI: 10.1111/cdev.13468

skill) a disability resides is inherently arbitrary, which is why prevalence estimates of dyslexia vary dramatically (see Francis et al., 2005; Hulme & Snowling, 2013). Fletcher (Fletcher, 2009; Fletcher & Miciak, 2017) has argued that since dyslexia does not represent a qualitatively distinct disorder, there is no need for separate theories of success and failure in learning to read, but instead greater focus on the normative development of reading across the entire distribution of reading skill. Similarly, Goswami (2003) has advocated for the need to better understand normative development when generating causal theories of developmental disorders such as dyslexia. She draws heavily on arguments by Karmiloff-Smith (1998) that disorders must be studied longitudinally to explore how alternative developmental pathways might lead to different phenotypical outcomes.

Arguments by Fletcher and Goswami notwithstanding, the majority of developmental dyslexia studies have tended to either prospectively or retrospectively divide the reading distribution into distinct dyslexic and typically developing groups. Both longitudinal group comparisons (e.g., de Jong & van der Leij, 2003; Peterson, Pennington, Olson, & Wadsworth, 2014) and reading-level match designs (e.g., Compton, 2002; Goswami et al., 2013; Metsala, 1999) have been used to compare children with dyslexia to those who are typically developing. While these designed comparisons have been crucial for establishing important, and potentially causal, cognitive models of dyslexia such as the phonological-core variable-difference model (Stanovich, 1988) and more contemporary multiple risk factor models (see Pennington, 2006; Pennington, & Bishop, 2009; Snowling & Hulme, 2012; Vellutino, Fletcher, Snowling, & Scanlon, 2004) they also present certain limitations (see Jackson & Butterfield, 1989). Both reading-level match and longitudinal group comparisons rely on the formation of dichotomous groups representing typically developing and dyslexic readers. There is considerable methodological literature demonstrating the negative consequences of the dichotomization of continuous variables (i.e., underestimating the strength of relations and reducing statistical power), firmly favoring the use of individual differences methods on continuous variables (e.g., MacCallum, Zhang, Preacher, & Rucker, 2002). In addition, variation in sampling procedures used to form groups of dyslexic and typically developing subjects can lead to sample-specific results that limit generalization and replication across studies (see Ramus, Altarelli, Jednoróg, Zhao, & di Covella, 2018).

Present Study

In this study, we were interested in examining the relationship between early growth (Grades 1–4) in WR and NWR skills in children with dyslexia, with the express intent of linking results to current behavioral and computational developmental frameworks of dyslexia (e.g., Compton, 2002; Harm & Seidenberg, 1999, respectively). To accomplish this, we adopted the approach advocated by Fletcher (2009) and Goswami (2003) by modeling parallel growth of WR and NWR factor scores across the entire distribution of reading skills using a growth curve of factors approach. Furthermore, vector plots were employed to visualize the relations between individual WR–NWR factor trajectories, referred to as codevelopment of WR and NWR, across Grades 1–4 as a function of initial WR and NWR skill in first grade; focusing especially on those children with low levels of initial reading skill in first grade (i.e., children at risk for developing dyslexia; in this study we consider children at the lower end of the first-grade WR and NWR distribution as being at high risk for developing dyslexia and therefore we use the term "dyslexic" to define this group. This label is used to link our results to the broader literature on developmental dyslexia but in no way implies these children have been diagnosed with developmental dyslexia.). These relatively new statistical modeling and data visualization techniques (Petscher, Quinn, & Wagner, 2016) allowed us to directly explore whether individual child-level WR–NWR factor trajectories align with the Orthography → Phonology (O → P) overfitting (Harm, McCandliss, & Seidenberg, 2003) and lexical asymmetry (Compton, 2002; Landerl, Frith, & Wimmer, 1996; Stanovich & Siegel, 1994) hypotheses of developmental dyslexia.

Aligning Computational and Behavioral Developmental Models of Dyslexia

Computational and behavioral models of WR have been proposed to explain the developmental relation between phonological processing, WR, and NWR (i.e., a measure of decoding skills) in typically developing and dyslexic children. We focus here on the connectionist model proposed by Harm and Seidenberg (Harm & Seidenberg, 1999, 2004; Harm et al., 2003) and the behavioral lexical asymmetry model proposed by various behavioral researchers (Compton, 2002; Landerl et al., 1996; Metsala, 1999; Stanovich & Siegel, 1994). While the computational and behavioral models of dyslexia vary in how

they conceptualize and ultimately operationalize the developing reading system, both models converge on a similar set of relations that develop between WR and NWR across time in typically developing and dyslexic children which are likely mediated by phonological processing skills. We briefly review the lexical asymmetry hypothesis (i.e., behavioral approach) and orthography → phonology (O → P) overfitting hypothesis (i.e., computational approach) describing the developmental relation between WR and NWR in typical children and children with dyslexia.

WR and Decoding Development

Acquiring a lexical system of representations that permits efficient word recognition is an essential part of learning to read (see Perfetti & Stafura, 2014). A typically developing reader's orthographic lexicon contains approximately 10,000 word-specific representations (excluding inflectional forms) by eighth grade (Ehri, 2005). This requires a lexical system that can quickly establish and reliably retrieve word-specific spellings (often referred to as word forms) that activate pronunciation and meaning. To accomplish this a child's orthographic lexicon changes in two important ways (e.g., Perfetti, 1992). First, there is a continuous increase in the absolute number of orthographically addressable entries, referred to as "word-specific" representations. Word-specific representations are considered to be *less* dependent on phonological processes because these representations have been supplanted by specific connections linking spelling directly to pronunciations (Perfetti, 1992; Share, 1995).

The second developmental change in the orthographic lexicon associated with reading development is an increase in the overall quantity and quality of subword O → P associations (Perfetti, 1992). Early O → P associations are initially based on simple one-to-one correspondences between graphemes and phonemes that are relatively insensitive to orthographic context (see Share, 1995). Over time with increased reading experience these subword connections develop into a set of abstract context-dependent relations between orthography and phonology that are "implicit, numerous, and very fast" (Gough, Juel, & Griffith, 1992). During reading development context-dependent subword associations evolve to represent the probabilistic co-occurrences and constraints that exist between orthographic and phonological units that resemble rule-like behavior (Perfetti, 1992; Seidenberg, 2005; Steacy et al., 2019). These subword associations are used by developing readers to decode unknown letter strings that allow new words to be added as word-specific representations in the evolving orthographic lexicon (see Share, 1995).

Thus, two lexical acquisition systems are at work as a child learns to read—the addition of word-specific entries and the expansion of subword O → P associations. These acquisition systems are mutually facilitative: "The more powerful the context-sensitive decoding rules (or analogic capabilities), the more entries the learner can acquire. And the more entries, the more powerful the decoding rules" (Perfetti, 1992, pp. 161–162). In typically developing readers these two systems develop together each supporting the continued expansion of the other. Disruptions in either of the two acquisition systems should compromise the system and affect reading development. We argue that asymmetry occurs when one of the lexical subsystems grows without concomitant growth in the other system, resulting in disruption of orthographic lexicon development and WR impairment.

Asymmetry Theory

Deficits in phonemic awareness skill likely limit the growth of important subword O → P associations in children with dyslexia (see Harm & Seidenberg, 2004). As a result, dyslexics exhibit a general tendency to process only partial information about words and further to rely on other sources of information that are considerably less efficient to facilitate word recognition (see Stanovich, 1988). This is consistent with the view that children with dyslexia may be over-reliant on a global processing strategy that results in insufficient attention to individual letters or groupings of letters (Ehri & Saltmarsh, 1995; Frith, 1985; Siegel, Share, & Geva, 1995) and the corresponding phonological representations. The result is that children with dyslexia tend to add word-specific representations to the lexicon without associated growth in subword O → P associations. From a grain size perspective (see Ziegler & Goswami, 2005), attention to the courser whole word → phonological grain prevents the development of finer O → P associations that support the development and expansion of important context-dependent subword O → P associations (Harm et al., 2003; Steacy et al., 2019). Various researchers (Compton, 2002; Compton, Miller, Elleman, & Steacy, 2014; Landerl et al., 1996; Metsala, 1999; Siegel et al., 1995) have conceptualized this as asymmetric development of the lexical system in which children with dyslexia add word-specific representations

(i.e., whole word → phonological grain) to their orthographic lexicons without the corresponding development of sublexical associations (i.e., fine-grained O → P associations) seen in typically developing readers. This asymmetry explanation is consistent with the behavioral observation of the NWR deficit in children with developmental dyslexia (e.g., Metsala et al., 1998; Rack et al., 1992).

Connectionist Models of Dyslexia

The research on the mechanisms of reading has given rise to a number of sophisticated computational models (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). Here we adopt the Triangle model of reading based on principles of the connectionist or PDP framework (Harm & Seidenberg, 1999, 2004; Seidenberg & McClelland, 1989). In the Triangle model, the reading system is instantiated in a network of neuron-like processing units that communicate by sending excitatory and inhibitory signals to one another. The network is organized into distinct "layers" (sets of nodes) including layers responsible for representing the various linguistic properties (orthographic, phonological, semantic) of a word, as well as layers of "hidden units" that function to mediate the associative mappings between the orthographic, phonological, and semantic layers. Like neural synapses, the connections in a network are plastic, and a learning algorithm is used to adjust their strengths (or weights) based on the interaction of the network and its task environment. Seeing a word causes a pattern of activation over the orthographic-phonological-semantic layers specific to that word. This results in a cascading flow of excitatory and inhibitory signals throughout the network, and over time the network settles into a stable pattern of activation that serves as the representation of that word.

The hidden units increase the computational power of the network by extending the range and complexity of problems the model can solve (Rumelhart, Hinton, & Williams, 1986). These hidden units have an interesting theoretical interpretation: they are the model's basis for developing underlying representations that abstract away from surface features of the input and output codes. Thus learning in such systems is not merely the creation of associations between patterns. Instead, the hidden units allow generalizations to be learned by permitting the patterns of activation over the hidden units to represent the intermediate code formed by this abstraction process. The existence of these intermediate units gives the networks a different character than older approaches in which behavior was construed as simple stimulus–response associations or associative chains. As such, hidden units allow the model to abstract important context-dependent relations between orthography and phonology that represent the probabilistic co-occurrences and constraints that exist between orthographic and phonological units (Seidenberg, 2005).

Computational implementations of the model have simulated a variety of aspects of reading, including skilled adult reading, reading acquisition in typically developing and children with dyslexia, and the impact of reading intervention (see Rueckl, 2016; Rueckl & Seidenberg, 2009; Seidenberg, 2005, for reviews). Because the connectionist architecture is, by design, distributed it is not wholly appropriate to describe separate "word-specific" and "subword" O → P associations as in the behavior models. Instead, distributed models allow O → P associations of all sizes to be represented in the connections between layers and these connections over time represent the natural co-occurrence of O → P associations in the reading system.

Connectionist models that incorporate moderate to severe phonological impairments have demonstrated a type of asymmetry in how O → P associations are represented in the model (for details see Harm & Seidenberg, 1999, 2004; Harm et al., 2003). All phonologically impaired models demonstrate impairments in NWR, with NWR performance being inversely correlated with the degree of phonological impairment. Mild phonological impairment only affects NWR while "sparing" WR, whereas more severe phonological impairments affect WR as well. In an unimpaired model, the phonological system contributes to decoding by "cleaning up" noisy or incomplete phonological representations. This allows the hidden units to provide the phonological system with somewhat incomplete input that captures the regularities in the mapping from orthography to phonology but can be easily overridden by the clean-up process in the case of exception words (e.g., *head*). In the impaired phonological system, more work has to be done by the hidden units because the phonological system is less able to repair degraded or incomplete phonological representations that result from the decoding process. This increased workload on the hidden units causes them to "memorize" word forms and form "item-specific" representations. Harm and Seidenberg hypothesize that the effect of requiring the hidden units to perform a more exact computation results in overfitting of the training data, which interferes with generalization (i.e.,

O → P overfitting hypothesis). As such, the impaired network treats words differently from each other, representing them more like unanalyzed, individual wholes with less overlapping structure. The impaired model cannot take advantage of the similarity between words when reading the nonword, even though it can correctly pronounce real words. The model provides a computational account of why poor phonological representations lead to poor reading, and in particular poor generalization of fine-grained O → P associations needed to read nonwords. Instead of forming representations sensitive to subword O → P units, the hidden units in the impaired simulations tend to learn item-specific representations. The formation of these item-specific representations is what directly impairs NWR. The poor NWR in the model is not due to the phonological system's impaired ability to assemble phonemes produced by the reading system, but rather the phonological impairment causes poor O → P representations to be formed during learning.

Convergence of Computational and Behavior Models

We see a convergence between the hypothesized time course of nonword and WR development according to computational and behavioral theories of dyslexia. Namely, initially weak phonological representations lead to a global processing strategy in word learning resulting in asymmetric development of the lexical system in children with dyslexia such that word-specific representations are added to the orthographic lexicon without the corresponding development of sublexical O → P associations that allow the decoding of nonwords. Here, we test this hypothesis by modeling parallel growth of WR and NWR factor scores across the entire distribution of reading skills using a growth curve of factors approach, focusing especially on those children with low levels of initial reading skill in first grade (i.e., children at risk for developing dyslexia). Furthermore, vector plots were used to visualize the codevelopment between individual WR–NWR factor trajectories across Grades 1–4 as a function of initial WR and NWR skill in first grade. This is the first study to model, and display visually, the relations between WR and NWR growth trajectories in a large sample of developing readers; thus, allowing the "asymmetry"/"O → P overfitting" hypotheses of dyslexia to be appraised. Based on the overfitting and asymmetry hypothesis, we posit that the WR–NWR trajectories of children who initially struggle to learn to read will show growth in WR skill without concomitant increases in NWR growth. Per Harm and Seidenberg (1999), we further suspect that those children who are at the lowest end of the distribution in initial WR will not only show a lack of NWR growth but also restricted growth in WR.

Method

Participants

Participants (N = 588) were drawn from a multiyear cohort longitudinal study examining response-to-intervention decision rules and prevention efficacy in first-grade children (see Compton et al., 2010; Gilbert et al., 2013). The sample represents the demographics of the local district with respect to the percent of African American (45.60% sample/47% district) and Caucasian (38.69% sample/35% district) children; 51.5% female, and 7.1 years (SD = 0.20 years) at the first wave of assessment. The sample had a lower percentage of Hispanic children compared to the district (6.22% sample/16% district) due to the initial sampling requirement across the three cohorts that children enrolled in first-grade English Language Learner instruction be eliminated from the sample. The mean full-scale IQ score for the sample was 98.20 (SD = 15.35). 61.47% of the children in the sample were receiving free or reduced lunch. For this study, children were assessed at the end of first through fourth grades on measures of timed and untimed WR and NWR skill. Additionally, we report the results of a phonemic awareness assessment administered in the fall of first grade.

Procedure

Test examiners were graduate research assistants who had been trained on tests until procedures were implemented with 90% fidelity. Students' WR and NWR were tested late in the spring semester of Grades 1–4 in a single session lasting less than 1 hr. Phonemic awareness skill was assessed as part of a language and cognition battery administered in the fall of first grade. All tests were given individually, audio recorded for reliability/fidelity purposes, and scored by the original examiner. All tests were double-scored and double-entered; discrepancies were resolved by a third examiner. Average fidelity of test administration procedures (based on a random selection of 20% of the taped assessment sessions) exceeded 95% for all tests.

Measures

Untimed WID Skill

Word identification was measured with the WID subtest from the Woodcock Reading Mastery Tests—Revised/Normative Update (Woodcock, 1998). For this task, children were asked to read words aloud one at a time. The test was not timed but children were encouraged to move to the next item after a 5-s silence. This measure was administered at the end of Grades 1–4. The examiner's manual reports split-half reliability exceeding 0.85 for the grade range of our sample (Woodcock, 1998).

Timed WID Skill

The Test of Sight Word Reading Efficiency (TOWRE: SWE; Torgesen, Wagner, & Rashotte, 1997) is a norm-referenced measure of sight WR accuracy and fluency. Children were presented with a list of 104 words ordered in ascending difficulty and asked to read aloud as many as possible in 45 s. This measure was administered at the end of Grades 1–4. Test–retest reliability exceeds 0.85 for the sample grade range (Torgesen et al., 1997).

Untimed NWR Skill

Nonword decoding was measured using the Woodcock Word Attack (WA) subtest from the Woodcock Reading Mastery Tests—Revised/Normative Update (Woodcock, 1998). Children were asked to read isolated pseudowords aloud (e.g., ree, ip, and weaf) one at a time. The test was not timed but children were encouraged to move to the next item after a 5-s silence. This measure was administered at the end of Grades 1–4. The manual reports split-half reliability exceeding 0.85 for the sample grade range (Woodcock, 1998).

Timed NWR Skills

The Test of Phonemic Decoding Efficiency (TOWRE: PDE; Torgesen et al., 1997) is a norm-referenced measure of decoding accuracy and fluency. For this test, children were presented with a list of 63 nonsense words ordered in ascending difficulty and asked to read (decode) aloud as many as possible in 45 s. This measure was administered at the end of Grades 1–4. Test–retest reliability exceeded 0.80 for the sample.

Phonological Awareness

Phonological awareness was measured with the Elision subtest of the Comprehensive Test of Phonological Processing (Wagner, Torgesen, & Rashotte, 1999). Children were presented a word, asked to repeat the word, and then asked to say the word without a specified syllable for the first three items and without a specified phoneme for the remaining 17 items. Items are ordered by increasing difficulty, and three consecutive incorrect items served as the ceiling. This measure was administered at the start of Grade 1. Coefficient alpha provided by the manual exceeds .85 for first-grade children (Wagner et al., 1999).

Analyses

Growth Models

To address the research question regarding the growth patterns of WR and NWR over time, we conducted a series of growth models using the MPlus software (Muthén & Muthén, 1998–2012). Rather than use first-order latent growth models, we chose to use the second-order curve of factors growth models. Although first-order latent growth models assess change in an unobserved construct from a single score, the second-order curve of factors growth models characterize the relation between multiple measures and their underlying construct across time. Here two measures (timed and untimed) were used to characterize WR and NWR; recognizing that dyslexia in opaque orthographies such as English is characterized by both inaccurate and slow WR and NWR performance (see Ziegler & Goswami, 2005). While the curve of factors approach has been around for some time (Isiordia & Ferrer, 2018; McArdle, 1988) it has rarely been applied in the social sciences. The curve of factors approach specifically capitalizes on the advantages inherent in modeling a construct's change over time with multiple measures.

The overall model is presented in Figure 1. Before fitting the parallel growth curve of factors model, we first modeled growth for WR and NWR separately. All models were based on raw score performance and were centered at the spring of first grade with factor loadings constrained to be equal across time (i.e., metric invariance with loadings constrained but intercepts and residual variances freely estimated). A particular goal in our work was to employ statistical techniques that allow for a visualization and understanding of the simultaneous, codevelopment of latent WR and NWR ability scores. The four waves of collected data allow for linear and quadratic functional forms to be tested in understanding WR and NWR trajectories.

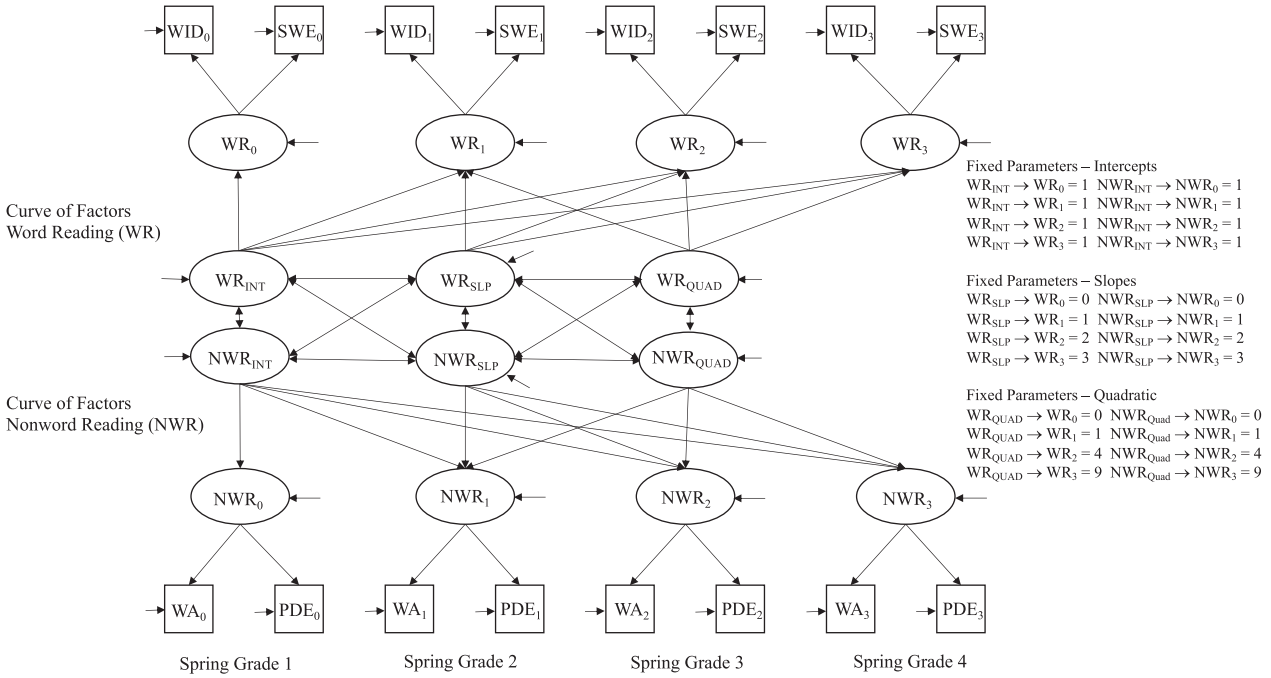

Figure 1. Final curve of factors model for WR and NWR. WR = word reading; NWR = nonword reading; WID = word identification; SWE = sight word efficiency; WA = Word Attack; PDE = phoneme decoding efficiency; INT = intercept; SLP = slope; QUAD = quadratic. 0–3 represents assessment wave.

Figure 2 displays plots of the estimated and observed data for the linear (top) and quadratic (bottom) models of NWR (left) and WR (right), clearly indicating a curvilinear form of growth across times, with deceleration, in both WR and NWR.

What primarily informed our modeling process was the extent to which the estimated trajectories could be easily understood and explained via visualizations of the codevelopment processes (see Vector Plots section). We were interested in modeling growth as accurately as possible to allow precise estimation of individual WR and NWR factor scores at the initial intercept (spring Grade 1) and final endpoint (spring Grade 4) for use in the vector plots. We evaluated the ability of both linear and quadratic models to fit the growth trajectories of WR and NWR across time individually and in a combined parallel growth model (see Bollen & Curran, 2006). The exact model building process is outlined below.

Vector Plots

To aid data visualization, unconditional vector plots for pair-wise codevelopment of WR and NWR were developed based on individual output from the growth models. Vectors are quantities that have both direction and magnitude; vector plots allow visualization of the direction and magnitude of simultaneous growth in two sets of constructs, but are not representative of the growth function of each construct separately (i.e., quadratic). In our case, the vector plot for estimated WR and NWR codevelopment (shown in Figure 3) displays estimated WR scores on the *x*-axis and estimated NWR scores on the *y*-axis. The two axes reflect the full range of estimated factor scores that capture development. Each arrow represents the WR–NWR vector for an individual child. The arrows provide several key pieces of information for interpreting codevelopment. First, the base of the arrow indicates the estimated initial status of a child's WR and NWR (i.e., spring of first grade). Second, the head of an arrow represents the final estimated factor scores on both the WR and NWR assessments. Third, the length of an arrow describes the amount of growth in both WR and NWR during Grades 1–4. Finally, the angle of the arrow characterizes codevelopment. Arrows pointed closer to 0° reflect individual growth that is dominant for the measured domain on the *x*-axis (i.e., WR), whereas arrows pointed closer to 45° represent relatively equal codevelopment in both WR and NWR domains (Petscher et al., 2016). Thus, each vector plot represents initial WR and NWR performance in first

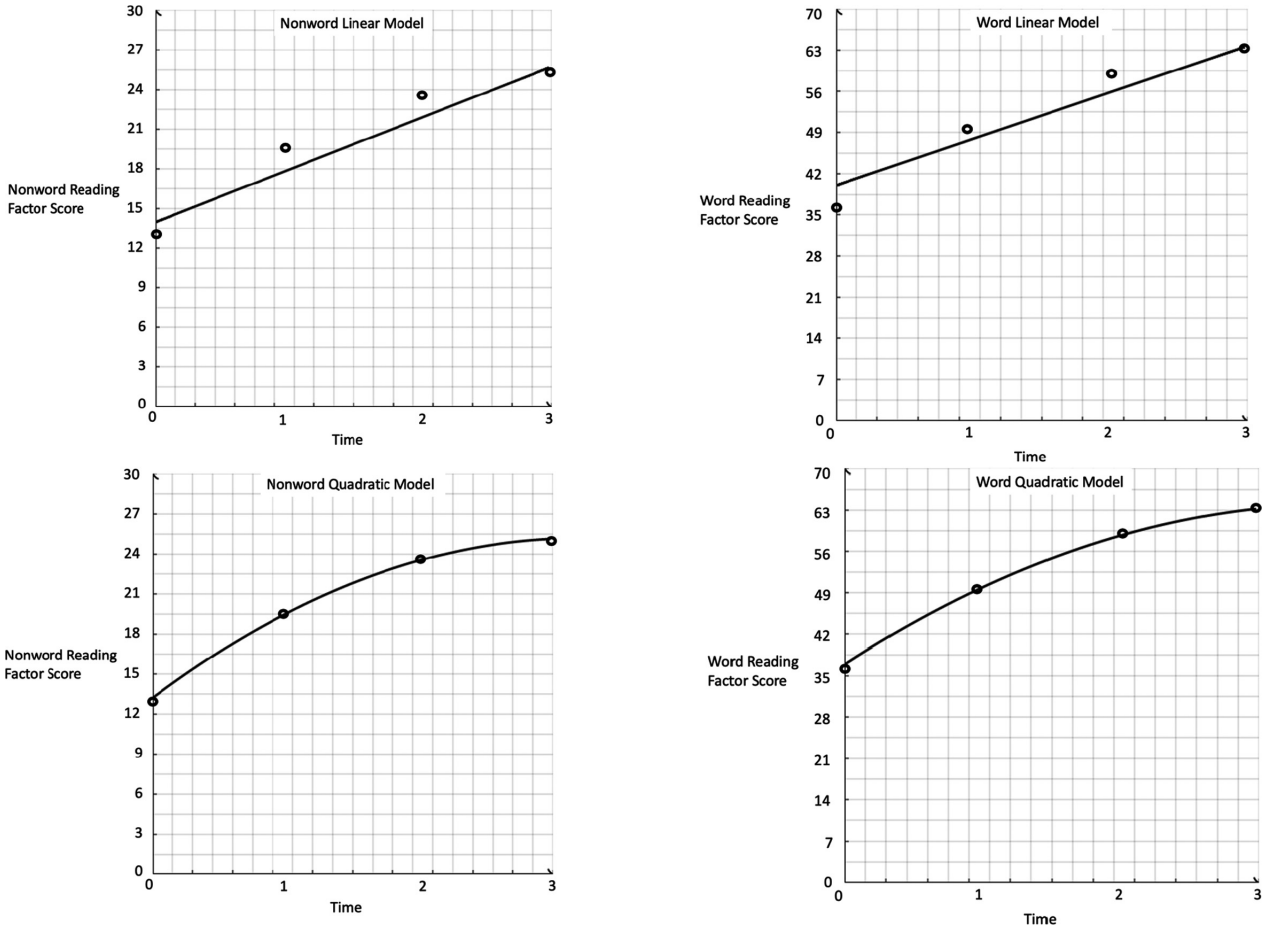

Figure 2. Plots of the estimated and observed data for the linear (top) and quadratic (bottom) models of nonword reading (left) and word reading (right) growth.

grade, final WR and NWR performance in fourth grade, the magnitude of growth in WR and NWR across Grades 1–4, and the relative codevelopment of WR and NWR growth between Grades 1 and 4.

Results

Zero-order correlations among the measures along with raw scores and norm-referenced performance on measures of WR, NWR, and phonemic awareness across the assessment waves are presented in Table 1. In general, norm-referenced performance across time indicates a relatively normally distributed sample with standard scores hovering around the national mean. As expected, all the correlations were significant ($p < .001$), with correlations between reading measures ranging from .62 to .94 and correlations between phonemic awareness and reading measures ranging from .53 to .64. Missing data were missing completely at random according to a Little's test ($\chi^2(78) = 81.93$, $p = .358$) and were estimated using maximum likelihood.

Determining the Factor Structure and Measurement Invariance

To determine the longitudinal structure of the WR and NWR factors, a series of confirmatory factor analyses were fit for WR and NWR at each grade level using the untimed measures of WID and WA (Woodcock, 1998) and timed measures of SWE and PDE (Torgesen et al., 1997). Chi-square difference tests were used to determine if a one- or two-factor model (i.e., WR and NWR) best fit the data at each time point. A two-factor model with a factor for WR (i.e., WID and SWE) and NWR (i.e., WA and PDE) fit the data significantly better than a one-factor model at each time point: Grade 1: $\Delta\chi^2(1) = 94.18$, $p < .01$; Grade 2: $\Delta\chi^2(1) = 58.41$, $p < .01$; Grade 3: $\Delta\chi^2(1) = 48.84$, $p < .01$; and Grade 4: $\Delta\chi^2(1) = 21.46$, $p < .01$. A similar procedure was

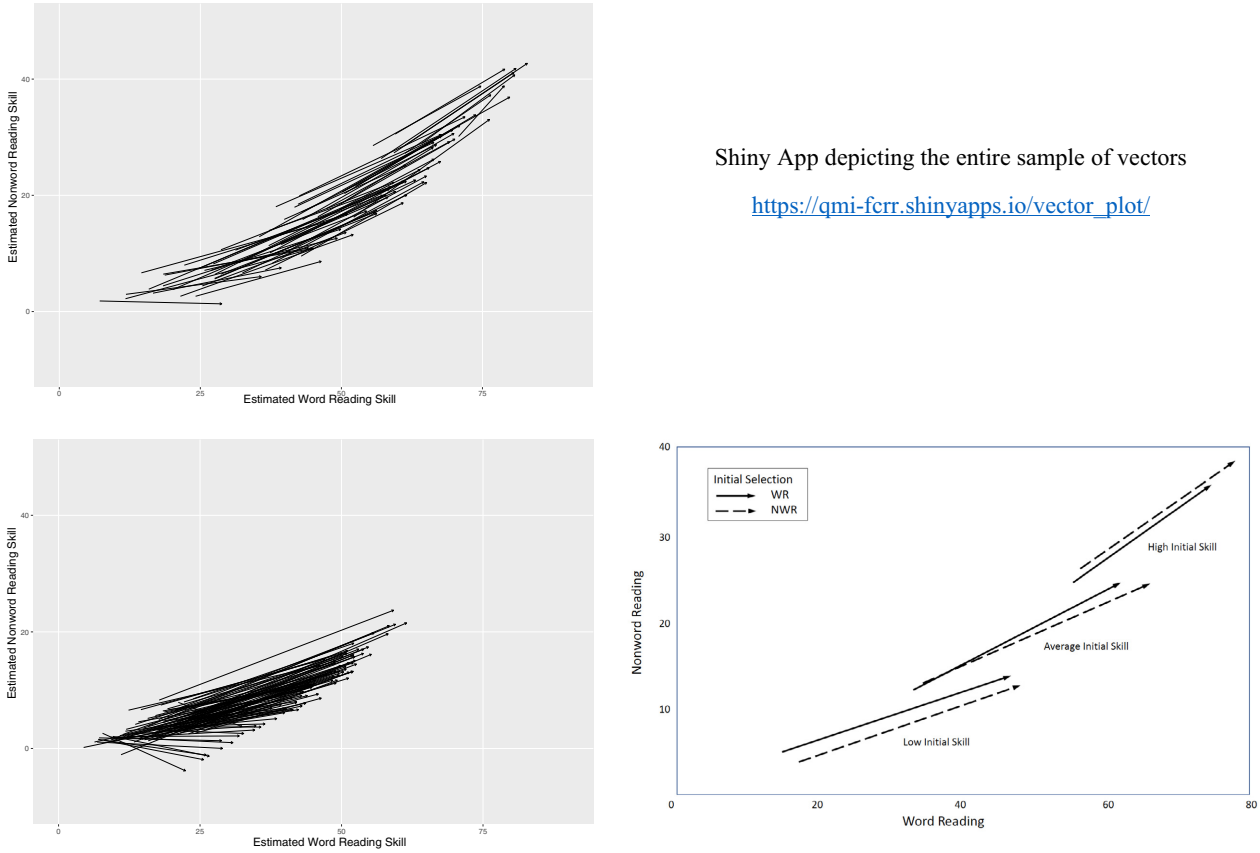

Figure 3. Vector plots illustrating estimated growth trajectories of word (*x*-axis) versus nonword (*y*-axis) reading (Grades 1–4) as a function of estimated initial word reading (WR) and nonword skill in first grade, in clockwise order starting in the upper left hand corner, for (a) random sampling of 25% of the subjects: (b) Shiny App depicting vectors as a function of first grade phonological awareness skill; (c); aggregated vectors for low, average, and high groups based on initial word and nonword reading (NWR) skill, and (d) total sample of subjects with low initial WR skill.

used to test longitudinal measurement invariance allowing factor loadings, factor intercepts, and residual variances to vary freely or be constrained to be equal across time. A metric invariance measurement model (i.e., loadings constrained across time but intercepts and residual variances freely estimated) was favored for WR ($\Delta\chi^2(3) = 6.12$, $p > .05$) and NWR ($\Delta\chi^2(3) = 5.62$, $p > .05$) growth over a more relaxed model allowing factor loadings to be estimated freely across time.

Univariate Growth in WR and NWR

Word Reading

To explore singular growth in WR, we first fit linear and quadratic growth models to the WR data using latent factors for WR across time (see top of Figure 1 with associated fixed growth parameter for intercept and slope values). The model that best fit the data was the quadratic model that allowed the intercept, slope, and quadratic parameters to correlate; WID error variances to correlate across time; and SWE error variances to correlate between grades three and four. The right panels of Figure 2 plot the estimated and observed WR factor data for linear (top) and quadratic (bottom) models of growth. Fit statistics and model comparisons for the linear and quadratic models; factor correlations for the quadratic model; and parameter estimates for the quadratic model are provided in Tables 2–4, respectively. Model fit statistics clearly favor the quadratic over the linear model for WR with fit indices generally meeting requirements proposed by Hu and Bentler, (1999). The chi-squared difference tests indicated that the addition of a quadratic term significantly improved model fit. Thus a quadratic model of growth was adopted to approximate growth over time in WR. In this model, the correlation between intercept and slope terms was −.35; intercept and quadratic .21; and slope and quadratic −.97. The mean intercept was 36.6 (factor score),

Table 1
Zero-Order Correlations Among Measures Along With Central Tendency Data

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. WID_0	—																
2. WID_1	.89	—															
3. WID_2	.86	.93	—														
4. WID_3	.83	.92	.94	—													
5. SWE_0	.93	.84	.80	.78	—												
6. SWE_1	.85	.91	.87	.86	.86	—											
7. SWE_2	.79	.86	.87	.86	.79	.91	—										
8. SWE_3	.76	.83	.83	.84	.76	.89	.91	—									
9. WA_0	.84	.76	.74	.71	.81	.71	.65	.62	—								
10. WA_1	.82	.86	.83	.81	.78	.79	.76	.70	.81	—							
11. WA_2	.78	.85	.86	.85	.74	.80	.78	.73	.74	.85	—						
12. WA_3	.77	.85	.85	.88	.73	.81	.79	.77	.72	.84	.90	—					
13. PDE_0	.83	.77	.75	.72	.85	.73	.68	.65	.86	.80	.75	.73	—				
14. PDE_1	.81	.85	.82	.81	.82	.84	.78	.75	.80	.86	.85	.83	.85	—			
15. PDE_2	.78	.83	.83	.83	.80	.83	.83	.78	.75	.83	.88	.85	.81	.90	—		
16. PDE_3	.76	.83	.84	.84	.77	.84	.84	.82	.73	.80	.87	.88	.78	.88	.93	—	
17. PA_{Fall}	.62	.63	.64	.61	.59	.59	.57	.53	.58	.61	.62	.59	.58	.59	.59	.57	—
RS Mean	36.7	50.9	58.9	63.8	34.3	49.3	57.2	63.0	13.3	20.3	23.6	25.4	11.1	20.0	25.0	28.5	12.9
SD	14.2	13.3	13.3	13.0	15.9	15.6	14.8	14.2	9.4	10.6	10.3	9.9	9.4	12.6	13.5	13.8	5.2
SS Mean	104.9	99.7	98.3	96.16	106.9	103.7	102.4	100.9	103.0	104.3	102.3	100.8	99.9	101.6	98.9	99.2	9.5
SD	11.9	12.4	12.9	12.41	12.0	14.2	13.9	13.4	15.1	16.1	14.5	13.8	11.4	14.5	15.7	15.7	2.2

Note. All correlations are significant at $p < .001$. 0–3 subscripts represent time points in years (Grades 1–4). WR = word reading; NWR = nonword reading; WID = word identification; SWE = sight word efficiency; WA = word attack; PDE = phonemic decoding efficiency; PA = phonemic awareness; RS = raw score; SS = standard scores ($M = 100$; $SD = 15$) or scaled scores ($M = 10$; $SD = 3$).

Table 2
Model Fit Statistics and Scaled Chi-Squared Difference Tests

	CFI	TFI	RMSEA	SRMR	*df*	χ^2	$\Delta\chi^2$
WR only							
Linear	.932	.874	.224	.079	15	459.02	
Quadratic	.994	.985	.077	.028	11	49.12	409.90*
NWR only							
Linear	.933	.866	.212	.084	14	385.35	
Quadratic	.985	.957	.121	.056	10	95.92	289.42*
Parallel growth							
Linear	.926	.902	.138	.069	91	1,110.44	
Quadratic	.975	.962	.086	.052	78	415.77	694.67*

Note. CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root mean square error of approximation; SRMR = standard root mean squared residual; $\Delta\chi^2 = \chi^2$ difference; WR = word reading; NWR = nonword reading.
*$p < .0001$

mean slope was 15.2 (factor score), and mean quadratic was −2.2 (factor score).

Nonword Reading

To explore singular growth in NWR, we again fit linear and quadratic growth models to the NWR data using latent factors for NWR across time (see bottom of Figure 1 with associated fixed growth parameter estimates for intercept and slope values). The model that best fit the data was the quadratic model that allowed the intercept and slope to correlate; allowed the WA error variances to correlate across time; allowed PDE error variances to

Table 3
Factor Correlations for the Parallel Growth Quadratic Model

Factor	WR_{INT}	WR_{SLP}	WR_{QUAD}	NWR_{INT}	NWR_{SLP}	NWR_{QUAD}
WR_{INT}	—					
WR_{SLP}	−.34	—				
WR_{QUAD}	.21	−.97	—			
NWR_{INT}	.92	.36	.21	—		
NWR_{SLP}	.43	.71	−.86	.36	—	
NWR_{QUAD}	−.52	−.60	.84	−.48	−.96	—

Note. All correlations are statistically significant, $p < .001$. WR = word reading; NWR = nonword reading; INT = intercept; SLP = slope; QUAD = quadratic.

Table 4
Factor Loadings and Means for the WR Only, NWR Only, and Parallel Growth Quadratic Models

	WR only model			NWR only model			Parallel model		
Parameters	EST	*SE*	*p*	EST	*SE*	*p*	EST	*SE*	*p*
WR factors[a]									
$WR_0 \rightarrow WID_0$	1.00	.00	.000				1.00	.00	.000
$WR_0 \rightarrow SWE_0$	1.13	.02	.000				1.14	.01	.000
$WR_1 \rightarrow WID_1$	1.00	.00	.000				1.00	.00	.000
$WR_1 \rightarrow SWE_1$	1.13	.02	.000				1.14	.01	.000
$WR_2 \rightarrow WID_2$	1.00	.00	.000				1.00	.00	.000
$WR_2 \rightarrow SWE_2$	1.13	.02	.000				1.14	.01	.000
$WR_3 \rightarrow WID_3$	1.00	.00	.000				1.00	.00	.000
$WR_3 \rightarrow SWE_3$	1.13	.02	.000				1.14	.01	.000
NWR factors[a]									
$NWR_0 \rightarrow WA_0$				1.00	.00	.000	1.00	.00	.000
$NWR_0 \rightarrow PDE_0$				1.29	.02	.000	1.14	.01	.000
$NWR_1 \rightarrow WA_1$				1.00	.00	.000	1.00	.00	.000
$NWR_1 \rightarrow PDE_1$				1.29	.02	.000	1.14	.01	.000
$NWR_2 \rightarrow WA_2$				1.00	.00	.000	1.00	.00	.000
$NWR_2 \rightarrow PDE_2$				1.29	.02	.000	1.14	.01	.000
$NWR_3 \rightarrow WA_3$				1.00	.00	.000	1.00	.00	.000
$NWR_3 \rightarrow PDE_3$				1.29	.02	.000	1.14	.01	.000
Means									
WR_{INT}	36.6	.60	.000				36.6	.60	.000
WR_{SLP}	15.2	.34	.000				15.2	.34	.000
WR_{QUAD}	−2.2	.09	.000				−2.2	.09	.000
NWR_{INT}				13.2	.37	.000	13.2	.37	.000
NWR_{SLP}				7.4	.31	.000	7.4	.31	.000
NWR_{QUAD}				−1.2	.90	.000	−1.2	.90	.000

Note. WR = word reading; NWR = nonword reading; WID = word identification; SWE = sight word efficiency; WA = word attack; PDE = phonemic decoding efficiency; INT = intercept; SLP = slope; QUAD = quadratic.
[a]Unstandardized estimates with invariant factor loadings across time.

correlate between grades three and four; and allowed the PDE and WA error variances to correlate between grades one and three, grades two and four, and grades three and four. The left panels of Figure 2 plot the estimated and observed NWR factor data for linear (top) and quadratic (bottom) models of growth. Fit statistics and model comparisons for the linear and quadratic models; factor correlations for the quadratic model; and parameter estimates for the quadratic model are provided in Tables 2–4, respectively. Model fit statistics clearly favor the quadratic over the linear model for NWR

with fit indices (excluding root mean square error of approximation [RMSEA]) generally meeting requirements proposed by Hu and Bentler, (1999). The chi-squared difference tests indicated that the addition of a quadratic term significantly improved model fit. Thus a quadratic model of growth was adopted to approximate growth over time in NWR. In this model, the correlation between intercept and slope terms was .36; intercept and quadratic −.48; and slope and quadratic −.96. The mean intercept was a 13.2 (factor score), mean slope was 7.4 (factor score), and mean quadratic was −1.2 (factor score).

Parallel Growth Model

After we modeled the singular growth of each skill, we modeled the parallel growth of WR and NWR, using a linear and quadratic model for each (see Figure 2). Again we compared differences in fit between parallel models based on linear and quadratic growth functions. Fit statistics and model comparisons for the linear and quadratic models; factor correlations for the quadratic model; and parameter estimates for the quadratic model are provided in Tables 2–4, respectively. Model fit statistics clearly favor the quadratic over the linear model for parallel growth with fit indices (excluding RMSEA) generally meeting requirements proposed by Hu and Bentler, (1999). The chi-squared difference tests indicated that the addition of a quadratic term significantly improved model fit. Thus a quadratic model of growth was adopted to approximate parallel growth of WR and NWR over time. In this model, the correlations between WR and NWR intercept (.92), slope (.71), and quadratic (.84) terms were highly significant.

Vector Plots Representing WR–NWR Trajectories Across the Distribution

Finally, we provide a series of vector plots in Figure 3 to simultaneously explore differential growth in WR and NWR across the full distribution of initial WR skills. Starting in the upper left-hand corner and moving clockwise, the first panel provides the vectors for a random sampling of 25% children in the sample. As mentioned earlier, these vector plots are visualizations that allow one to view the initial starting skill in WR and NWR, final estimated skill in both WR and NWR, relative amount of growth in WR and NWR, and the simultaneous growth relations between WR and NWR (i.e., angle) across the entire distribution of children. The first panel indicates almost 45° growth in the highest skilled initial readers, with a steady decrease in the degree of the vectors as initial skill decreases. At the lower end of the distribution, the vectors become flat suggesting that there are students who are demonstrating WR growth without any concomitant growth in NWR.

The next panel in Figure 3 provides a link to a Shiny App allowing visualization of growth (https://qmi-fcrr.shinyapps.io/vector_plot/) as a function of the fall of first-grade phonemic awareness skill (indicated by color gradients). The results suggest that students who begin the study with poor phonological awareness have a higher tendency to make fewer gains in NWR than WR across time and have less overall growth. However, the results indicate that initially poor phonemic awareness skill is not sufficient to fully explain children who show relatively horizontal vectors, suggesting there may be other factors that explain individual differences in vector slopes in the lower range of initial reading skill.

The third panel in Figure 3 aggregates vectors across students who initially start low (lowest 16% of the sample on estimated WR), average (middle 68% of the sample on estimated WR), and high (highest 16% of the sample on estimated WR) in WR and NWR skill. As can be seen, the general trends seen between WR–NWR vectors and initial performance in WR hold at the group level with vector slopes decreasing as group initial WR skill decreases. As predicted by both the Orthography → Phonology overfitting and lexical asymmetry hypotheses the three groups differed significantly on phonemic awareness skill assessed in the fall of first grade ($F(3,585) = 711.80$, $p < .001$; post hoc comparisons indicating low ($M = 6.86$, $SD = 2.40$) < average ($M = 9.02$, $SD = 2.27$) < high ($M = 11.66$, $SD = 2.62$), indicating that the low initial WR group also struggled with the phonemic awareness tasks assessed in first grade (M = Norm-referenced standard scores with a mean of 10 and standard deviation of 3). As would be expected phonemic awareness skill measured in the fall of first grade correlated significantly with the spring first grade factor score representing WR ($r = .61$, $p < .001$) and NWR ($r = .63$, $p < .001$).

The final panel shows the entire distribution of vectors for children whose initial WR skill was 1 *SD* below the group mean. This figure demonstrates a wide variation in the magnitude of WR growth, NWR growth, and WR–NWR codevelopment in the lower end of the sample with a large number of children exhibiting horizontal vectors. Children with the lowest initial WR and horizontal

NWR also appear to have the lowest magnitude of WR growth. However, it is clear that large individual differences exist in the relations between WR and NWR in the group of children with the lowest initial WR skills.

Discussion

In this study, we examined parallel growth of WR and NWR factor scores in a large developmental sample of students from first to fourth grade using a growth curve of factors approach along with vector plots to visualize WR–NWR codevelopment across the entire distribution of initial WR skills. Visualization of the relation between initial WR and WR–NWR codevelopment offers unique insights into the relations between WR and NWR development in children who struggle initially to learn to read. We interpret our results within a broad set of theories, based upon computational and behavioral approaches, suggesting that children with dyslexia overfit O → P associations across time, relying on more "global" processing strategies, resulting in lexical asymmetry in which word-specific representations are added to the lexicon without concomitant addition of sublexical O → P associations needed to read nonwords. We begin our discussion by linking our results with previous studies examining dyslexia symptomatology associated with WR and NWR development. We then move on to speculate on mechanistic factors that could lead to the range of individual differences in WR–NWR trajectories at the lower end of the initial WR skill distribution. Finally, we conclude with a discussion of implications for early identification and treatment and future considerations.

Dyslexia Symptomatology During Early Reading Development

In general, vector visualization indicates that children with low initial WR and NWR skills tend to exhibit relatively flat WR–NWR codevelopment (i.e., WR growth without concomitant NWR growth) supporting the well-documented NWR deficit in children with developmental dyslexia (see Rack et al., 1992). Additionally, results indicate that while WR and NWR growth factors are highly related (r = .71) in the sample, the relation between WR and NWR trajectories changes in a nonlinear fashion as a function of initial WR and NWR skill. Specifically, the angle of WR–NWR vectors move from nearly 45° in children with average and high initial WR skill to much more horizontal in children with low initial WR. Flat vectors indicate WR growth without concomitant growth in NWR. In addition, children with low initial WR skills also had lower phonological awareness skills compared to children with average and high initial WR skills.

From both a connectionist modeling and behavioral asymmetry perspective, this failure to generalize WR skills to reading novel nonwords suggests that children with dyslexia may be over-reliant on a global processing strategy that affords insufficient attention to individual letters or groupings of letters (Ehri & Saltmarsh, 1995; Frith, 1985; Siegel et al., 1995) and the corresponding phonological representations (Harm & Seidenberg, 1999, 2004). Thus, the observed nonlinear relation between initial WR and the WR–NWR codevelopment vectors are consistent with explanations offered by both the Triangle model and the lexical asymmetry models of developmental dyslexia.

Mechanistic Factors Effecting WR–NWR trajectories in Dyslexic Children

Previous studies suggesting that dyslexic readers show greater sensitivity to orthographic structure (i.e., "superior" orthographic skills) have been taken as indicating that dyslexics use a more visually based strategy in reading (see Siegel et al., 1995; Stanovich & Siegel, 1994). However, our data favor an alternative explanation in which dependence on whole-word orthographic structure is not a preferred strategy but instead a consequence of how the mapping between orthography and phonology is learned when the capacity to represent phonological structure is limited. From a lexical asymmetry standpoint, Ehri and Saltmarsh (1995) have hypothesized that children with dyslexia lack sufficient "grapho-phonic" knowledge to fully analyze matches between orthographic and phonological units in order to store high-quality representations that include complete sublexical O → P associations. This is consistent with the view that phonological processing deficits in children with dyslexia leads to a processing strategy that affords insufficient attention to sublexical O → P relations (e.g., Compton, 2002; Compton et al., 2014; Landerl et al., 1996; Metsala, 1999).

In terms of the Triangle model's overfitting hypothesis, poor NWR is caused by the formation of "noncomponential" O → P associations that are more holistic mappings (i.e., overfitting results in exact O → P associations that limit generalization). Componential representations share structure with

other items, and hence pronunciation can be aided through this overlapping structure; noncomponential representations, on the other hand, lead to poor generalization because the overlap in structure is less apparent (see Harm & Seidenberg, 1999). In the Triangle model, such noncomponential representations are themselves caused by poor phonological representations at the beginning of the acquisition of reading. Given that learning is dependent on existing representations; once poor representations have been formed, subsequent learning exploits the (poor) characteristics of these existing representations (Harm et al., 2003). It is important to emphasize that both the lexical asymmetry and the overfitting hypotheses view this dependence on a more global orthographic structure in dyslexia not as a strategy but as a consequence of how O → P mappings are learned.

Implication for Early Identification and Treatment

Evidence that poor readers form word-specific representations that do not generalize can be used to provide insights into diagnostic tests that determine the locus of WR problems and suggest interventions tuned to the specific source of the impairment. Certainly, our results suggest that contrasting early WR and NWR skill growth across time in struggling readers could help to quickly identify those children with characteristic flat WR–NWR trajectories that indicate elevated risk for poor word-level reading outcomes. Furthermore, the instruction that children are exposed to plays a role in how they learn to read words and how this affects the decoding of novel words (and nonwords), particularly in children struggling to learn to read. Instructional routines that focus students' attention on connections at the sublexical level in a systematic way have an impact on the pattern of results observed regarding real-WR versus decoding. Thompson, Connelly, Fletcher-Flinn, and Hodson (2009) observed differences in adult decoding patterns depending on whether participants had received phonics or no phonics in childhood reading instruction. They found that adults who had received phonics instruction read words that had alternative acceptable pronunciations using more regular, context-free grapheme-phoneme correspondences than adults who had not received phonics instruction as children, suggesting an apparent "cognitive footprint" of childhood phonics instruction.

Both behavioral and computational approaches seem to suggest that systematically building sublexical O → P associations is key to remediating poor generalization of subword regularities. Behaviorally, McCandliss, Beck, Sandak, and Perfetti (2003) explored an intervention that targeted the development of sublexical connections during WR. At pretest, at-risk students were generally cuing off of the first letter in words, with little focus on internal sounds of words. After receiving an intervention that targeted subword letter-sound correspondences (using a technique that builds each new word by changing a single letter in the previous word), students performed significantly better on decoding and WR outcomes. They theorized that this intervention placed pressure on the O → P system to form mappings more sensitive to the internal parts or components of words, perhaps altering the initial cognitive footprint that favored the global processing of words. Similarly, Blachman et al. (2014) employed an intervention focused on building sublexical O → P associations in a study evaluating the long term effects of an 8-month intervention on second and third graders (selected on the basis of poor word-level skills) reading with significant WR effects favoring the treatment surviving a decade later (d = .62).

Simulations of intervention effects using the Triangle model (Harm et al., 2003) have demonstrated that phonological awareness remediations that ignore the O → P mapping system result in limited improvements, particularly after poor representations have already been formed in the O → P system because they failed to place direct pressure on the system to change mappings that were already learned. Whereas, interventions that explicitly focus on promoting sublexical O → P mappings, making them more componential, had a greater impact on model performance, especially for generalization to novel words. Furthermore, Harm and Seidenberg (1999) comment that, "the introduction of a training regime that actively discourages the computation of the sound pattern of a word could result in poor generalization performance in the absence of phonological impairments" (p. 512). Suggesting the importance of the initial "cognitive footprint" in determining whether WR and decoding codevelop.

Future Considerations

Nation and Castles (2017) note in their review of orthographic learning that processes other than phonological decoding may help to explain individual differences in how children develop an efficient lexical system of representations that permits efficient word recognition. They note that is not clear

what other cognitive skills, beyond phonological decoding, play a causal role in orthographic learning. Certainly, within the self-teaching framework, a child's ability to discover and eventually exploit the statistical regularities between orthography and phonology represented in the lexicon varies as a function of a child's ability to induce statistical regularities (a form of statistical learning), the corpus of words the child has been exposed to, and the strength of the underlying knowledge structures that facilitate item-level learning (see Steacy, Elleman, & Compton, 2017). In addition to the importance of establishing sublexical connections in reading, evidence suggests that students require the acquisition of novel phonological representations for successful reading. The work of Litt and colleagues (Litt, de Jong, van Bergen, & Nation, 2013; Litt & Nation, 2014) suggests that students require phonological representations of words (i.e., familiarity with words) prior to reading. Through a series of paired associative learning (PAL) tasks, they found that there was an impact of PAL on reading only in conditions when students were required to say the (new, monosyllabic) names corresponding to what they saw or heard. Results suggest that it is the ability to form novel phonological representations, as revealed by the verbal output demand, that counts in the relation between PAL and reading. More importantly, the lower performance of the children with dyslexia that they observed in the subsequent visual-verbal PAL task could be fully accounted for by their preceding verbal learning performance—confirming a deficit in phonological form learning (Litt & Nation, 2014). The researchers have also demonstrated that strong phonological representations and the ability to flexibly relate decoded forms to established phonological representations predict WR. Referred to as a person's "set for variability," Elbro, de Jong, Houter, and Nielsen (2012) offer that such an association is done by means of a kind of attractor network (as proposed by Harm & Seidenberg, 1999). In such a network, existing phonological representations (the sounds of known words) work like individual magnets on incoming spoken words. Hence, incoming words will be attracted to (and activate) the word in the mental lexicon that has the closest resemblance—regardless of whether words are heard or generated by print. A combination of these approaches to exploring the development of the lexical system may afford us a better understanding of the individual differences observed in this study and may provide further insight into the codevelopment of WR and NWR in children who initially struggle to learn to read. Finally, we recognize that our curve of factor modeling approach employed the minimum manifest variables needed for estimation. Future models should include more manifest measures of WR and NWR so that the latent variables are more precisely estimated.

References

Blachman, B. A., Schatschneider, C., Fletcher, J. M., Murray, M. S., Munger, K. A., & Vaughn, M. G. (2014). Intensive reading remediation in grade 2 or 3: Are there effects a decade later? *Journal of Educational Psychology*, *106*(1), 46. https://doi.org/10.1037/a0033663

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective* (Vol. *467*). Hoboken, NJ: Wiley.

Bruck, M. (1993). Word recognition and component phonological processing skills of adults with childhood diagnosis of dyslexia. *Developmental Review*, *13*, 258–268. https://doi.org/10.1006/drev.1993.1012

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, *108*, 204–256. https://doi.org/10.1037//0033-295X.108.1.204

Compton, D. L. (2002). The relationship between phonological processing, orthographic processing, and lexical development in reading-disabled children. *Journal of Special Education*, *35*, 201–210. https://doi.org/10.1177/002246690203500402

Compton, D. L., Miller, A. C., Elleman, A. M., & Steacy, L. M. (2014). Have we forsaken reading theory in the name of "quick fix" interventions for children with reading disability?. *Scientific Studies of Reading*, *18*(1), 55–73.

Elbro, C., de Jong, P. F., Houter, D., & Nielsen, A. M. (2012). From spelling pronunciation to lexical access: A second step in word decoding?. *Scientific Studies of Reading*, *16*(4), 341–359.

Compton, D. L., Fuchs, D., Fuchs, L. S., Bouton, B., Gilbert, J. K., Barquero, L. A., . . . Crouch, R. C. (2010). Selecting at-risk first-grade readers for early intervention: Eliminating false positives and exploring the promise of a two-stage gated screening process. *Journal of Educational Psychology*, *102*, 327. https://doi.org/10.1037/a0018448

de Jong, P. F., & van der Leij, A. (2003). Developmental changes in the manifestation of a phonological deficit in dyslexic children learning to read a regular orthography. *Journal of Educational Psychology*, *95*(1), 22. https://doi.org/10.1037/0022-0663.95.1.22

Ehri, L. C. (2005). Learning to read words: Theory, findings, and issues. *Scientific Studies of Reading*, *9*, 167–188. https://doi.org/10.1207/s1532799xssr0902_4

Ehri, L. C., & Saltmarsh, J. (1995). Beginning readers outperform older disabled readers in learning to read

words by sight. *Reading and Writing*, *7*, 295–326. https://doi.org/10.1007/BF03162082

Fletcher, J. M. (2009). Dyslexia: The evolution of a scientific concept. *Journal of the International Neuropsychological Society*, *15*, 501–508. https://doi.org/10.1017/S1355617709090900

Fletcher, J. M., & Miciak, J. (2017). Comprehensive cognitive assessments are not necessary for the identification and treatment of learning disabilities. *Archives of Clinical Neuropsychology*, *32*, 2–7. https://doi.org/10.1093/arclin/acw103

Francis, D. J., Fletcher, J. M., Stuebing, K. K., Lyon, G. R., Shaywitz, B. A., & Shaywitz, S. E. (2005). Psychometric approaches to the identification of LD: IQ and achievement scores are not sufficient. *Journal of Learning Disabilities*, *38*, 98–108. https://doi.org/10.1177/00222194050380020101

Frith, U. (1985). Beneath the surface of developmental dyslexia. In K. Patterson, J. Marshall, & M. Coltheart (Eds.), *Surface dyslexia* (pp. 301–330). London. UK: Erlbaum.

Gilbert, J. K., Compton, D. L., Fuchs, D., Fuchs, L. S., Bouton, B., Barquero, L. A., & Cho, E. (2013). Efficacy of a first-grade responsiveness-to-intervention prevention model for struggling readers. *Reading Research Quarterly*, *48*, 135–154. https://doi.org/10.1002/rrq.45

Goswami, U. (2003). Why theories about developmental dyslexia require developmental designs. *Trends in Cognitive Sciences*, *7*, 534–540. https://doi.org/10.1016/j.tics.2003.10.003

Goswami, U., Mead, N., Fosker, T., Huss, M., Barnes, L., & Leong, V. (2013). Impaired perception of syllable stress in children with dyslexia: A longitudinal study. *Journal of Memory and Language*, *69*(1), 1–17. https://doi.org/10.1016/j.jml.2013.03.001

Gough, P. B., Juel, C., & Griffith, P. L. (1992). Reading, spelling, and the orthographic cipher. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 35–48). Hillsdale, NJ: Erlbaum.

Harm, M. W., McCandliss, B. D., & Seidenberg, M. S. (2003). Modeling the successes and failures of interventions for disabled readers. *Scientific Studies of Reading*, *7*, 155–182. https://doi.org/10.1207/S1532799XSSR0702_3

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*, *106*, 491–528. https://doi.org/10.1037/0033-295X.106.3.491

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review*, *111*, 662–720. https://doi.org/10.1037/0033-295X.111.3.662

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55. https://doi.org/10.1080/10705519909540118

Hulme, C., & Snowling, M. J. (2013). *Developmental disorders of language learning and cognition*. Hoboken, NJ: Wiley.

Isiordia, M., & Ferrer, E. (2018). Curve of factors model: A latent growth modeling approach for educational research. *Educational and Psychological Measurement*, *78*, 203–231. https://doi.org/10.1177/0013164416677143

Jackson, N. E., & Butterfield, E. C. (1989). Reading-level-match designs: Myths and realities. *Journal of Reading Behavior*, *21*, 387–412. https://doi.org/10.1080/10862968909547686

Karmiloff-Smith, A. (1998). Development itself is the key to understanding developmental disorders. *Trends Cognitive Science*, *2*, 389–398. https://doi.org/10.1016/S1364-6613(98)01230-3

Landerl, K., Frith, U., & Wimmer, H. (1996). Intrusion of orthographic knowledge on phoneme awareness: Strong in normal readers, weak in dyslexic readers. *Applied Psycholinguistics*, *17*(1), 1–14. https://doi.org/10.1017/S0142716400009437

Litt, R. A., de Jong, P. F., van Bergen, E., & Nation, K. (2013). Dissociating crossmodal and verbal demands in paired associate learning (PAL): What drives the PAL–reading relationship?. *Journal of Experimental Child Psychology*, *115*(1), 137–149.

Litt, R. A., & Nation, K. (2014). The nature and specificity of paired associate learning deficits in children with dyslexia. *Journal of Memory and Language*, *71*(1), 71–88.

Lovett, M. W., Borden, S. L., DeLuca, T., Lacerenza, L., Benson, N. J., & Brackstone, D. (1994). Testing the core deficits of developmental dyslexia: Evidence of transfer of learning after phonologically- and strategy-based reading training programs. *Developmental Psychology*, *30*, 805–822. https://doi.org/10.1037/0012-1649.30.6.805

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7*(1), 19. https://doi.org/10.1037/1082-989X.7.1.19

McArdle, J. J. (1988). Dynamic but structural equation modeling of repeated measures data. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (pp. 561–614). Boston, MA: Springer.

McCandliss, B., Beck, I. L., Sandak, R., & Perfetti, C. (2003). Focusing attention on decoding for children with poor reading skills: Design and preliminary tests of the word building intervention. *Scientific Studies of Reading*, *7*(1), 75–104. https://doi.org/10.1207/S1532799XSSR0701_05

Metsala, J. L. (1999). The development of phonemic awareness in reading-disabled children. *Applied Psycholinguistics*, *20*(1), 149–158. https://doi.org/10.1017/S0142716499001058

Metsala, J. L., Stanovich, K. E., & Brown, G. D. A. (1998). Regularity effects and the phonological deficit model of reading disabilities: A meta-analytic review. *Journal of Educational Psychology*, *90*, 279–293. https://doi.org/10.1037/0022-0663.90.2.279

Muthén, L. K., & Muthén, B. O. (1998–2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Author.

Nation, K., & Castles, A. (2017). Putting the learning into orthographic learning. In K. Cain, D. L. Compton, & R. K. Parrila (Eds.), *Theories of reading development. Vol. 15: Studies in written language and literacy* (pp. 147–168). Amsterdam, The Netherlands: John Benjamins.

Pennington, B. F. (2006). From single to multiple deficit models of developmental disorders. *Cognition*, *101*, 385–413. https://doi.org/10.1016/j.cognition.2006.04.008

Pennington, B. F. (2009). *Diagnosing learning disorders: A neuropsychological framework* (2nd ed.). New York, NY: Guilford Press.

Pennington, B. F., & Bishop, D. (2009). Relations among speech, language, and reading disorders. *Annual Review of Psychology*, *60*, 283–306. https://doi.org/10.1146/annurev.psych.60.110707.163548

Perfetti, C. A. (1992). The representation problems in reading acquisition. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 145–174). Hillsdale, NJ: Erlbaum.

Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, *18*(1), 22–37. https://doi.org/10.1080/10888438.2013.827687

Peterson, R. L., & Pennington, B. F. (2015). Developmental dyslexia. *Annual Review of Clinical Psychology*, *11*, 283–307. https://doi.org/10.1146/annurev-clinpsy-032814-112842

Peterson, R. L., Pennington, B. F., Olson, R. K., & Wadsworth, S. J. (2014). Longitudinal stability of phonological and surface subtypes of developmental dyslexia. *Scientific Studies of Reading*, *18*, 347–362. https://doi.org/10.1080/10888438.2014.904870

Petscher, Y., Quinn, J. M., & Wagner, R. K. (2016). Modeling the co-development of correlated processes with longitudinal and cross-construct effects. *Developmental Psychology*, *52*, 1690. https://doi.org/10.1037/dev0000172

Rack, J. P., Snowling, M. J., & Olson, R. K. (1992). The nonword reading deficit in developmental dyslexia: A review. *Reading Research Quarterly*, *27*(1), 29–53. https://doi.org/10.2307/747832

Ramus, F., Altarelli, I., Jednoróg, K., Zhao, J., & di Covella, L. S. (2018). Neuroanatomy of developmental dyslexia: Pitfalls and promise. *Neuroscience and Biobehavioral Reviews*, *84*, 434–452. https://doi.org/10.1016/j.neubiorev.2017.08.001

Rueckl, J. G. (2016). Toward a theory of variation in the organization of the word reading system. *Scientific Studies of Reading*, *20*(1), 86–97. https://doi.org/10.1080/10888438.2015.1103741

Rueckl, J. G., & Seidenberg, M. S. (2009). Computational modeling and the neural bases of reading and reading disorders. In R. Frost (Ed.), *How children learn to read: Current issues and new directions in the integration of cognition, neurobiology and genetics of reading and dyslexia research and practice* (pp. 101–133). New York, NY: Psychology Press.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. *1*, pp. 318–362). Cambridge, MA: MIT Press.

Seidenberg, M. S. (2005). Connectionist models of word reading. *Current Directions in Psychological Science*, *14*, 238–242. https://doi.org/10.1111/j.0963-7214.2005.00372.x

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*, 523. https://doi.org/10.1037/0033-295X.96.4.523

Share, D. L. (1995). Phonological recoding and self-teaching: Sine qua non of reading acquisition. *Cognition*, *55*, 151–218. https://doi.org/10.1016/0010-0277(94)00645-2

Shaywitz, S. E., Escobar, M. D., Shaywitz, B. A., Fletcher, J. M., & Makuch, R. (1992). Evidence that dyslexia may represent the lower tail of a normal distribution of reading ability. *New England Journal of Medicine*, *326*, 145–150. https://doi.org/10.1056/NEJM199201163260301

Shaywitz, S. E., Fletcher, J. M., Holahan, J. M., Shneider, A. E., Marchione, K. E., Stuebing, K. K., . . . Shaywitz, B. A. (1999). Persistence of dyslexia: The Connecticut longitudinal study at adolescence. *Pediatrics*, *104*, 1351–1359. https://doi.org/10.1542/peds.104.6.1351

Siegel, L. S., Share, D., & Geva, E. (1995). Evidence for superior orthographic skills in dyslexics. *Psychological Science*, *6*, 250–254. https://doi.org/10.1111/j.1467-9280.1995.tb00601.x

Snowling, M. J., & Hulme, C. (2012). Annual research review: The nature and classification of reading disorders—A commentary on proposals for DSM-5. *Journal of Child Psychology and Psychiatry*, *53*, 593–607. https://doi.org/10.1111/j.1469-7610.2011.02495.x

Stanovich, K. E. (1988). Explaining the differences between the dyslexic and the garden-variety poor reader: The phonological-core variable-difference model. *Journal of Learning Disabilities*, *21*, 590–604. https://doi.org/10.1177/002221948802101003

Stanovich, K. E., & Siegel, L. S. (1994). Phenotypic performance profile of children with reading disabilities: A regression-based test of the phonological-core variable-difference model. *Journal of Educational Psychology*, *86*(1), 24. https://doi.org/10.1037/0022-0663.86.1.24

Steacy, L. M., Compton, D. L., Petscher, Y., Elliott, J. D., Smith, K., Rueckl, J., . . . Pugh, K. (2019). Development and prediction of context-dependent vowel pronunciation in elementary readers. *Scientific Studies of Reading*, *23*, 49–63. https://doi.org/10.1080/10888438.2018.1466303

Steacy, L. M., Elleman, A. M., & Compton, D. L. (2017). Opening the 'black box'of learning to read: Inductive mechanisms supporting word acquisition development with a focus on children who struggle to read. In K. Cain, D. L. Compton, & R. K. Parrila (Eds.), *Theories of*

reading development (pp. 99–121). Amsterdam, The Netherlands: John Benjamins.

Thompson, G. B., Connelly, V., Fletcher-Flinn, C. M., & Hodson, S. J. (2009). The nature of skilled adult reading varies with type of instruction in childhood. *Memory & Cognition*, *37*, 223–234. https://doi.org/10.3758/MC.37.2.223

Torgesen, J. K. (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research & Practice*, *15*(1), 55–64. https://doi.org/10.1207/SLDRP1501_6

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1997). *Test of Word Reading Efficiency*. Austin, TX: Pro-Ed.

Vellutino, F. R., Fletcher, J. M., Snowling, M. J., & Scanlon, D. M. (2004). Specific reading disability (dyslexia): What have we learned in the past four decades? *Journal of Child Psychology and Psychiatry*, *45*(1), 2–40.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive Test of Phonological Processing*. Austin, TX: Pro-Ed.

Woodcock, R. W. (1998). *Woodcock Reading Mastery Tests —Revised/Normative Update: Examiner's manual*. Circle Pines, MN: American Guidance Service.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychological Bulletin*, *131*(1), 3. https://doi.org/10.1037/0033-2909.131.1.3

This document is a scanned copy of a printed document. No warranty is given about the accuracy of the copy. Users should refer to the original published version of the material.